PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
G06F 17/28, 17/21, 17/30

A1

(11) International Publication Number: WO 00/14651

(43) International Publication Date: 16 March 2000 (16.03.00)

(21) International Application Number: PCT/US99/19699

(22) International Filing Date: 31 August 1999 (31.08.99)

(30) Priority Data:

60/099,641	9 September 1998 (09.09.98)	US
09/321,804	27 May 1999 (27.05.99)	US

(71) Applicant: INVENTION MACHINE CORPORATION [US/US]; 133 Portland Street, Boston, MA 02114-1722 (US).

(72) Inventors: TSOURIKOV, Valery, M.; 770 Boylston Street, Boston, MA 02199 (US). BATCHILO, Leonid, S.; 40 Agassiz Avenue, Belmont, MA 02478-1323 (US). SOVPEL, Igor V.; Apartment 193, 3/1 Voronyanskogo Street, Minsk, Belarus 220039 (BY).

(74) Agent: DREYFUS, Edward; Suite 350, 608 Sherwood Parkway, Mountainside, NJ 07092 (US).

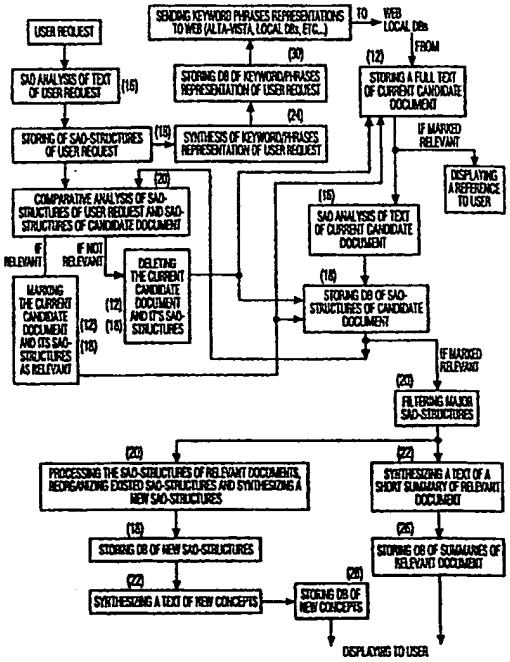
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: DOCUMENT SEMANTIC ANALYSIS/SELECTION WITH KNOWLEDGE CREATIVITY CAPABILITY

(57) Abstract

A computer based software system and method for semantically processing a user entered natural language request to identify (16) and store (18) linguistic subject–action–object (SAO) structures, using such structures as key words/phrases (24) to search (30) local and Web–based databases for downloading (12) candidate natural language documents, semantically processing candidate document texts into candidate document SAO structures, and selecting and storing only relevant documents whose SAO structures include a match with a stored request SAO structure. Further features include analyzing relationships among relevant document SAO structures and creating new SAO structures (20) based on such relationships that may yield new knowledge concepts and ideas for display to the user and generating and displaying natural language summaries (22, 26) based on the relevant document SAO structures.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

TITLE: DOCUMENT SEMANTIC ANALYSIS/SELECTION WITH KNOWLEDGE CREATIVITY CAPABILITY

REFERENCE TO PRIORITY APPLICATION:

This application claims the benefit of U.S. Provisional Application No. 60/099,641, filed September 9, 1998.

BACKGROUND:

The present invention relates to computer based apparatus for and methods of semantically analyzing, selecting, and summarizing candidate documents containing specific content or subject matter.

Computer based document search processors are known to perform key word searches for publications on the Internet and World Wide Web. Today, information owners and service providers are adapting their data bases to individual tastes and requirements. For example, Boston based Agents, Inc. offers over the Web personalized newsletters for music fans such that classical music lovers are blocked from receiving Rap music ads and vice-versa. KD, Inc. of Hong Kong has developed a system that takes into consideration words similar by sense while searching the Web. Today the user can download 10,000 papers from the Web by typing the word "Screen". The search system designed by KD, Inc. asks the user whether he/she is seeking papers related to Computer Screen, TV Screen or Window Screen. In this case, the number of unrelated

papers will be drastically reduced.

Software based search processors are able to remember requests of single user and to conduct personalized non-stop searches on the Web. So, when a user wakes up in the morning he/she finds references and abstracts of several new Web papers, related to his/her area of interest. In 1997, practically all fundamental technical publications, journals, magazines, as well as patents of all industrial countries became available on the Web, i.e. available in electronic format.

Although key word searching the Web affords the user great value, it also has created and will continue to create substantial problems adversely affecting this value. Specifically, because of the enormous amount of information available on the Web, key word search processors produce too much downloaded information, the vast majority of which is irrelevant or immaterial to the information the user wants. Many users simply give up in frustration when presented with several hundred articles in response to what the user considered a request for only those few articles related to a specific request.

This problem is also experienced in the technical fields of science and engineering, particularly since there is a growing number of libraries, government patent offices, universities, government research centers, and other adding vast amounts of technical and scientific information for Web access. Engineers, scientists, and doctors are overwhelmed with too many articles, papers, patents and general information on the topic of interest to them. In addition, the user presently has only two choices when examining a download article to determine its relevance to the users project. He/she can either read the authors abstract and/or scan various sections of the full article to determine whether or not to

save or print-out that specific document. Since the author's abstract is not comprehensive, it often omits the reference to the specific subject matter of interest to the user or treats this subject matter in an incomprehensive manner. Thus, scanning the abstract and scanning the full article may have little value and require an inordinate amount of user time.

Various attempts purport to increase the recall and precision of the selection such as U.S. Patents Nos. 5,774,833 and 5,794,050 incorporated here by reference, however, these methods simply rely on key word or phrase searching with various techniques of selection based on variations of the key words, or purported understanding of textual phrases. These prior methods may improve recall but may still requires too much physical and mental effort and time to determine why the document was selected and what is the pertinent part. This results from the entire document of abstract being presented without summary or concept generation.

SUMMARY OF EXEMPLARY EMBODIMENT OF PRESENT INVENTION

A computer based software system and method according to the principles of the present invention solves the foregoing problems and has the ability to perform a non-stop search of all databases on the Web or other network for key words and to semantically process candidate documents for specific technological functions and specific physical effects so that only the very few prioritized or a single article meeting the search criteria is presented or identified to the user.

Further, the computer based software system in accordance with the principles of

the present invention captures these few highly relevant documents and creates a compressed, short summary of the precise technical physical aspects designated by the search criteria.

Another aspect of the present invention includes using the semantic analysis results of the selected documents to create new ideas of knowledge concepts. The system does this by analyzing the subjects, actions, and objects mentioned in the documents and re-organizing these representations into new and/or different profiles of such elements. As further described below, some of these reorganized sets of relationships among these elements may comprise new concepts never before thought of by anyone.

According to an aspect of the present invention, the method and apparatus begins with the user entering natural language text related to the task or concept for which the user desires to acquire publications or documents. The system analyzes this request text and automatically tags each word with a code that indicates the type of word it is. Once all words in the request are tagged, the system performs a semantic analysis that, in one example, includes determining and storing the verb groups within the first sentence of the request, then determining and storing the noun groups within that sentence of the request. This process is repeated for all sentences in the request.

Next, the system parses each request sentence with an heirarcal algorithm into a coded framework which is substantially indicative of the sense of the sentence. The system includes databases of various types to aid in generating the coded framework, such as grammar rules, parsing rules, dictionary synonyms, and the like. Once parsed sentence codes are stored, the system identifies Subject-Action-Object (SAO) extractions within

each sentence and stores them. A sentence can have one, two, or a plurality of SAO extractions as seen in the detailed description below. Each extraction is normalized into a SAO structure by processing extractions according to certain rules described below. Accordingly, the result of the semantic analysis routine performed on the request text is a series of SAO structures indicative of the content of the request. These request SAO structures are applied to (1) a comparative module for comparing the SAO structures of candidate documents as described below and (2) a search request and key word generator that identifies key words and key combinations of words, and synonyms thereof, for searching the Web internet, intranet, and local data bases for candidate documents. Any suitable search engine, e.g. Alta Vista, can be used to identify, select, and download candidate documents based on the generated key words.

It should be understood that, as mentioned above, key word searching produces an over-abundance of candidate documents. However, according to the principles of the present invention, the system performs substantially the same semantic analysis on each candidate document as performed on the user input search request. That is, the system generates an SAO structure(s) for each sentence of each candidate document and forwards them to the comparative Unit where the request SAO structures are compared to the candidate document SAO structures. Those few candidate documents having SAO structures that substantially match the request SAO structure profile are placed into a retrieved document Unit where they are ranked in order of relevance. The system then summarizes the essence of each retrieved document by synthesizing those SAO structures of the document that match the request SAO structures and stores this summary for user

display or printout. Users can later read the summary and decide to display or print out or delete the entire retrieved document and its SAO's.

As stated above, the SAO structures for each sentence for each retrieved document are stored in the system according to the present invention. According to the knowledge creativity aspect of the present invention, the system analyzes all these stored structures, identifies where common or equivalent subjects and objects exist and reorganizes, generates, synthesizes, new SAO structures or new strings of SAO structures for user's consideration. Some of these new structured or strings may be unique and comprise new solutions to problems related to the user's requested subject matter. For example, if two structures S1-A1-O1 and S2-A2-O2 are stored, and the present system recognizes that S2 is equivalent to or the synonym for or has some other relation to O1 then it will generate and store for the user's access a summary of S1-A1-S2-A2-O2. Of if the system stores an association between S1 and A2 it can generate S1-A1/A2-O1 to suggest improvement of O1 toward desired results.

Other and further advantages and benefits shall become apparent with the following detailed description when taken in view of the appended drawings, in which:

DRAWING DESCRIPTION:

Figure 1 is a pictorial representation of one exemplary embodiment of the system according to the principles of the present invention.

Figure 2 is a schematic representation of the main architectural elements of the system according to the present invention.

Figure 3 is a schematic representation of the method according to the principles of the present invention.

Figure 4 is a schematic representation of Unit 16 of Figure 2.

Figure 5 is a schematic representation of Unit 20 of Figure 2.

Figure 6 is a schematic representation of Unit 22 of Figure 2.

Figure 7 is a typical example of the user request text entered by user.

Figure 8 is a tagged and coded representation version of text of Figure 7.

Figure 9 is an identification of verb groups of the text of Figure 8.

Figure 10 is an identification of noun groups of the coded text of Figure 8.

Figure 11 is a representation of parsed hierarchy coded text of Figure 8.

Figure 12 is a representation of SAO extraction of the text of Figure 7.

Figure 13 is a representation of SAO structures of the extraction of Figure 12.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

One exemplary embodiment of a semantic processing system according to the principles of the present invention includes:

A CPU 12 that could comprise a general purpose personal computer or networked server or minicomputer with standard user input and output driver such as keyboard 14, mouse 16, scanner 19, CD reader 17, and printer 18. System 10 also includes standard communication ports 21 to LANs, WANs, and/or public or private switched networks to the Web.

With reference to Figures 1-6, the semantic procession system 10 includes a

temporary storage or data base 12 for receiving and storing documents downloaded from the Web or local area net or generated as a user request text with use of keyboard 14 or one of the other input devices. User can type the request, examples disclosed below, or enter full documents into DB 12 and designate the document as user's request. System 10 further includes semantic processor 14 for receiving the entire text of each document and includes a Subject-Action-Object (SOA) analyzer Unit 16 that tags each word of each sentence with a code type (such as Markov chain theory code). Unit 16 then identifies each verb group and noun group, (described below) within each sentence, and parses and normalizes each sentence into SAO structures that represents the sense of the sentence. Unit 16 applies its output to DB of SAO structures 18. SAO processor Unit 20 stores the request SAO structures and receives the SAO structures of each sentence of each document stored in Unit 18. Unit 20 compares the document SAO's to the request SAO's and deletes out those documents with no matches. The SAO structures of matched documents are stored back in Unit 18 or some other storage facility. In addition, Unit 20 analyzes SAO structures within a single document or with those of one or more other relevant documents, searches for relationships among S-A-O's and generates new SAO structures for user consideration. These new structures are stored in Unit 18 or some other storage facility in the system.

Unit 14 further includes natural language Unit 22 that receives SAO structures in table form and synthesizes structures into natural language form, i.e. sentences.

Unit 14 also includes keyword Unit 24 for receiving SAO structures and extracts key words and phrases from them and acquires their synonyms for use as additional key

words/phrases.

Database Units 26, 28, and 30 receive the outputs from Unit 14, generally as shown, for storing the natural language summaries of selected SAO structures as described below and the key words/phrases that form user request sent to search engines through port 21.

Unit 16 includes document pre-formatter 32 that receives full text of documents from Unit 12 and converts the text and other contents to a standard plain text format. Text coder 34 analyzes each word of each sentence of text and tags a code to every word which code designates the word type, see Fig 8. Various data bases designated 44 in Fig 4 are available to aid the Units of Unit 16. Following tagging, recognizer Unit 36 identifies the verb groups (Fig 9) and the noun groups of each sentence (Fig 10). Sentence parser 38 then parses each sentence into a hierarchical coded form that represents the sense of the sentence. Fig 11. S-A-O extractor 40 organizes the SAO's of each sentence into extracted table format (Fig 12). Then normalizer 42 normalizes the extractions into SAO structures as described above (Fig 13).

SAO processor 20 includes three main Units. Comparative Unit 46 receives SAO structures from database 18. One set of these structures originates from the user request text described above and other sets originate from the candidate documents. Unit 46 then compares these two sets looking for matches between SAO structures of these two sets. If no match results, then the candidate document and associated SAO's are deleted. If a match is identified then the document is marked relevant and ranked and stored in Unit 12 and its SAO structures stored in Unit 18. Unit 46 then compares all candidate documents

in sequence and in the same way as described.

Unit 20 also includes the SAO structure reorganizing Unit 48 to synthesize new SAO structures from different documents on the same matter and combines them into the new structure, as described above, and applies them to Unit 18.

Filtering Unit 50 analyzes every SAO structure of each document and blocks or deletes those not relevant to the SAO structures of the request.

Reference 52 designates some of the data bases available to aid sub-units of Unit 20.

SAO synthesizer Unit 22 (Figure 6) includes a Subject detector 54 for detecting the content of the subject for each received SAO structure. If S is detected then the SAO is fed to Unit 56 in which the tree structure of the verb group(s) is restored to natural language using grammar, semantic, speech patterns, and synonyms rules data base 66. Synthesizer 58 does the same for subject noun groups and synthesizer 60 does the same for object noun groups. Combiner 68 then organizes and combines these groups into a natural language sentence.

If S was not detected by Unit 54, the SAO structures are processed by synthesizer 62 to restore the verb group in passive form. Synthesizer 64 processes the object noun group for a passive sentence and combiner 70 to organize and combine the groups into a natural language sentence.

If SAO structures received by Unit 54 bear new structure markings, then combiners 68 and 70 apply their output to Unit 28 and if they were marked existing SAO structure, then units 68, 70 apply output to Unit 26. See Fig 3.

The salient steps to the method according to the principles of the present invention

are shown in Figure 3, where the number in the parenthesis refer to the Units of Figure 2 where the process steps takes place. A session begins with the user inputting a national language request which could be customized with the use of the keyboard or would be a national language document entered via one of the input devices shown in Figure 1. A typical user generates customized request as shown in Figure 7., System 10 Unit 14, then by first tagging each word with a type code (See Figure 8) then identifying the verb groups of each sentence (Figure 9) and noun groups of each sentence (Figure 10) then processing each sentence into an hierarchical tree (Figure 11) and then extracting the SAO extractions where all extracted words are the originals of the request (Figure 12). Then the method normalizes these words (modifies) each as each action is changed to its infinitive form. This, "is isolated" Figure 12 is changed to "ISOLATE", the word "to" being understood (Figure 13). It should be understood that not all attributes of the subject, action and objects appearing in Figure 11 are shown in Figures 12 and 13, but the system knows the full attributes associated with the SAO elements and these attributes are part of the SAO structure. Also, note in Figure 13, no subject is listed for the last action because is indicated pursuant to the planning rules. This absence does not affect the reliability of the overall method because all sentences of the candidate documents the include an A-O of Isolate-slides will be considered a matter regardless of the subject. The normalized SAO's are called herein as SAO structures. These users request SAO structures are stored and applied in two following steps (i) synthesis of key word/phrases of user request; (ii) a comparative analysis of SAO structure of each sentence of each candidate documents as described below.

The request SAO structure key words/phrases are stored and sent to a standard search engine to search for candidate documents in local databases, LANs and/or the Web. AltaVista™, Yahoo™, or other typical search engines could be used. The engine, using the request SAO structure key words/phrases identifies candidate documents and stores them (full text) for system 10 analysis. Next the SAO analysis as described above for the search request is repeated for each sentence of each candidate document so that SAO structures are generated and stored as indicated in Fig. 3. In addition, the SAO structures of each document are used in the comparative steps where the request SAO structures are compared with the candidate document SAO structures. If no match is found then the documents and related SAO structures are deleted from the system. If one or more matches are found then the document and related structures are marked relevant and its relevancy marked for example on a scale of 1.0 to 10.0. The full relevant document text is permanently stored (although it can later be deleted by user if desired) for display or print-out as user desires. Relevant SAO structures are also marked relevant and permanently stored.

Next System 10 filters out the least relevant SAO structures and uses the matched SAO structures of each relevant documents to synthesis into natural language summary sentences(s) the matched SAO structures and the page number where the complete sentence associated with the matched SAO structures appears. This summary is stored and available for users display or print-out as desired.

Filtered relevant SAO structures of relevant document(s) are analyzed to identify relationships among the subjects, actions, and objects among all relevant structures. Then

SAO structures are processed to reorganize them into new SAO structures for storage and synthesis into natural language new sentence(s). The new sentences may and probably some of them will express or summarize new ideas, concepts and thoughts for users to consider. The new sentences are stored for user display or print-out.

For example, if

S_1-A_1-O_1

S_2-A_2-O_2

S_3-A_3-O_3

and S_1 is the same as or a synonym of O_3 then S_3-A_3-S_1-A_1-O_1 is synthesized into a new sentence and stored.

Accordingly, the method and apparatus according to the present invention provides user automatically with a set of new ideas directly relating to user's requested area of interest some of which ideas are probably new and suggest possible new solutions to user's problems under consideration and/or the specific documents and summaries of pertinent parts of specific documents related directly to user's request.

Although mention has been made herein of application of the present system and method to the engineering, scientific and medical fields, the application thereof is not limited thereto. The present invention has utility for historians, philosophers, theology, poetry, the arts or any field where written language is used.

It will be understood that various enhancements and changes can be made to the example embodiments herein disclosed without departing from the spirit and scope of the present invention.

WE CLAIM:

Claim 1. A natural language document analysis and selection system comprising,

a general purpose computer having a monitor, a central processing unit (CPU), a user input device for generating request data representing a natural language request, and a communications device for communication with local and remote natural language document databases,

said CPU comprising (i) first storage means for storing the request data, (ii) a semantic processor for generating request subject-action-object (SAO) extractions in response to receiving request data, and (iii) SAO storage means for storing representations of the request SAO extractions.

Claim 2. A system as set forth in Claim 1, wherein said communication device conveys candidate document data to said CPU for storage in said first storage means, the candidate document data representing natural language document text,

said semantic processor generating candidate document SAO extractions in response to receiving candidate document data, and

said SAO storage means also storing representations of candidate document SAO extractions.

Claim 3. A system as set forth in Claim 2, wherein said semantic processor identifies matches between said representations of said request SAO extractions and said candidate document SAO extractions.

Claim 4. A system as set forth in Claim 3, wherein said semantic processor comprises means for marking as relevant candidate document data that includes at least one representation of candidate document SAO extraction that matches at least one representation of request SAO extraction.

Claim 5. A system as set forth in Claim 4, wherein said semantic processor comprises means for deleting stored candidate document data and stored representations of candidate document SAO extractions for those documents that have no representation of candidate document SAO extraction that matches a representation of request SAO extraction.

Claim 6. A system as set forth in Claim 3, wherein said semantic processor includes an SAO text analyzer having a plurality of stored text formatting rules, coding rules, word tagging rules, SAO recognizing rules, parsing rules, SAO extraction rules, and normalizing rules for applying such rules to the request data and candidate document data such that said representations of candidate document SAO extractions and of request SAO extractions comprise candidate document and request SAO structures, respectively.

Claim 7. A system as set forth in Claim 6 further comprising second storage means for storing request SAO structures and for applying SAO structures as key words/phrases to said communication device for application to document search engines on the WEB or local databases to cause downloading of candidate document data to the system.

Claim 8. A system as set forth in Claim 6 further comprising an SAO synthesizer for generating and storing for display on said monitor natural language summaries of marked documents in response to receipt of document SAO structures.

Claim 9. A system as set forth in Claim 6 further comprising an SAO synthesizer for analyzing relationships among subjects, actions, and objects among relevant and stored SAO structures and processing those SAO structures that have a relationship with at least one other SAO structure to generate a different SAO structure and storing the different SAO structure for display to the user.

Claim 10. A system as set forth in Claim 9 wherein said relationship comprises:

S_1-A_1-O_1

S_2-A_2-O_2

where S_1 synonym O_2

Then S_2-A_2-S_1-A_1-O_1 .

Claim 11. In a digital data processing system including the World Wide Web and a general purpose computer having a monitor, a central processing unit (CPU), a user input device, and a communications device for communication with local and remote natural language document databases, the method of analyzing and selecting natural language documents comprising,

generating request data representing a natural language request,

storing the request data,

semantically processing the request data to generate request subject-action-object (SAO) extractions, and

storing representations of the request SAO extractions.

Claim 12. The method as set forth in Claim 11, wherein said communication device conveys candidate document data to said CPU, the candidate document data representing natural language document text,

storing the candidate document data,

said semantically processing including generating candidate document SAO extractions in relation to the candidate document data, and

storing representations of candidate document SAO extractions.

Claim 13. A method as set forth in Claim 12, wherein said semantically processing includes identifying matches between said representations of said request SAO extractions and said candidate document SAO extractions.

Claim 14. A method as set forth in Claim 13, wherein said semantically processing comprises marking as relevant candidate document data that includes at least one representation of candidate document SAO extraction that matches at least one

representation of request SAO extraction.

Claim 15. A method as set forth in Claim 14, wherein said semantically processing comprises deleting access to stored candidate document data and stored representations of candidate document SAO extractions for those documents that have no representation of candidate document SAO extraction that matches a representation of request SAO extraction.

Claim 16. A method as set forth in Claim 13, wherein said semantically processing includes applying a plurality of stored text formatting rules, noun and verb recognition rules, coding rules, word tagging rules, SAO recognizing rules, parsing rules, SAO extraction rules, and normalizing rules to the request data and candidate document data such that said representations of candidate document SAO extractions and representations of request SAO extractions comprise candidate document and request SAO structures, respectively.

Claim 17. A method as set forth in Claim 16 further comprising storing request SAO structures and applying SAO structures as key words/phrases to document search engines on the WEB or local databases to cause downloading of candidate document data to the CPU.

Claim 18. A method as set forth in Claim 16 further comprising generating and storing and displaying on said monitor natural language summaries of marked relevant documents in relation to relevant document SAO structures.

Claim 19. A method as set forth in Claim 16 further comprising analyzing relationships among subjects, actions, and objects among relevant and stored SAO structures, further processing those SAO structures that have a relationship with at least one other relevant and stored SAO structure, and generating a different SAO structure based on the said relationship, and

storing the different SAO structure and displaying the different SAO structure to the user.

Claim 20. A method as set forth in Claim 19 wherein said relationship comprises:

S_1-A_1-O_1 comprises one relevant and stored SAO structure

S_2-A_2-O_2 comprises a second relevant and stored SAO structure

where said relationship comprises S_1 synonym O_2

and the different SAO structure is

S_2-A_2-S_1-A_1-O_1 .

Claim 21. A method as set forth in Claim 19 wherein said relationship comprises:

S_1-A_1-O_1 comprises one relevant and stored SAO structure

S_2-A_2-O_2 comprises a second relevant and stored SAO structure

where said relationship exists between S_1 and A_2

and the different SAO structure is

S_1-A_1/A_2-O_1

where / means alternate.

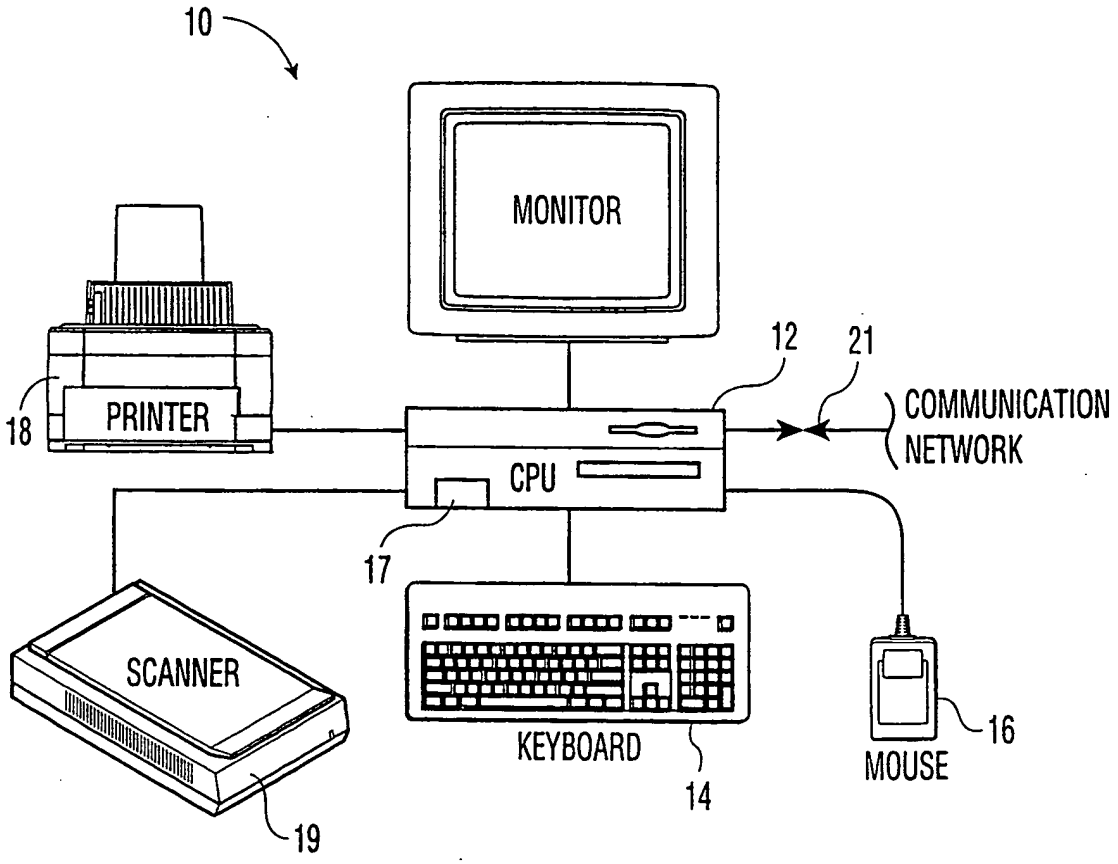

10
MONITOR
12
21
COMMUNICATION
NETWORK
18
PRINTER
CPU
17
SCANNER
KEYBOARD
14
16
MOUSE
19

FIG. 1

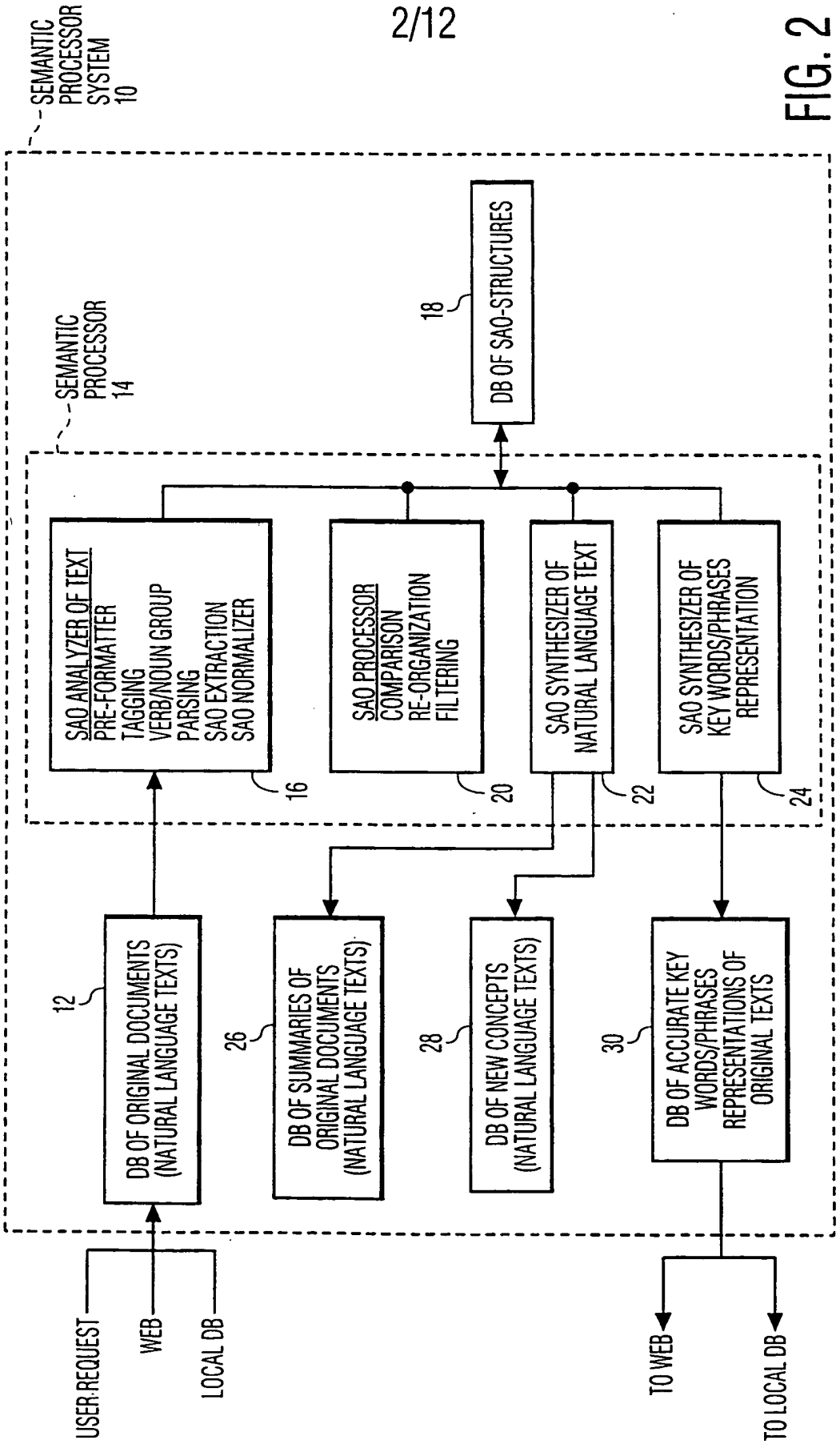

SEMANTIC PROCESSOR SYSTEM 10
SEMANTIC PROCESSOR 14
18
DB OF SAO-STRUCTURES
SAO ANALYZER OF TEXT
PRE-FORMATTER
TAGGING
VERB/NOUN GROUP
PARSING
SAO EXTRACTION
SAO NORMALIZER
16
SAO PROCESSOR
COMPARISON
RE-ORGANIZATION
FILTERING
20
SAO SYNTHESIZER OF NATURAL LANGUAGE TEXT
22
SAO SYNTHESIZER OF KEY WORDS/PHRASES REPRESENTATION
24
12
DB OF ORIGINAL DOCUMENTS (NATURAL LANGUAGE TEXTS)
26
DB OF SUMMARIES OF ORIGINAL DOCUMENTS (NATURAL LANGUAGE TEXTS)
28
DB OF NEW CONCEPTS (NATURAL LANGUAGE TEXTS)
30
DB OF ACCURATE KEY WORDS/PHRASES REPRESENTATIONS OF ORIGINAL TEXTS
USER REQUEST
WEB
LOCAL DB
TO WEB
TO LOCAL DB

FIG. 2

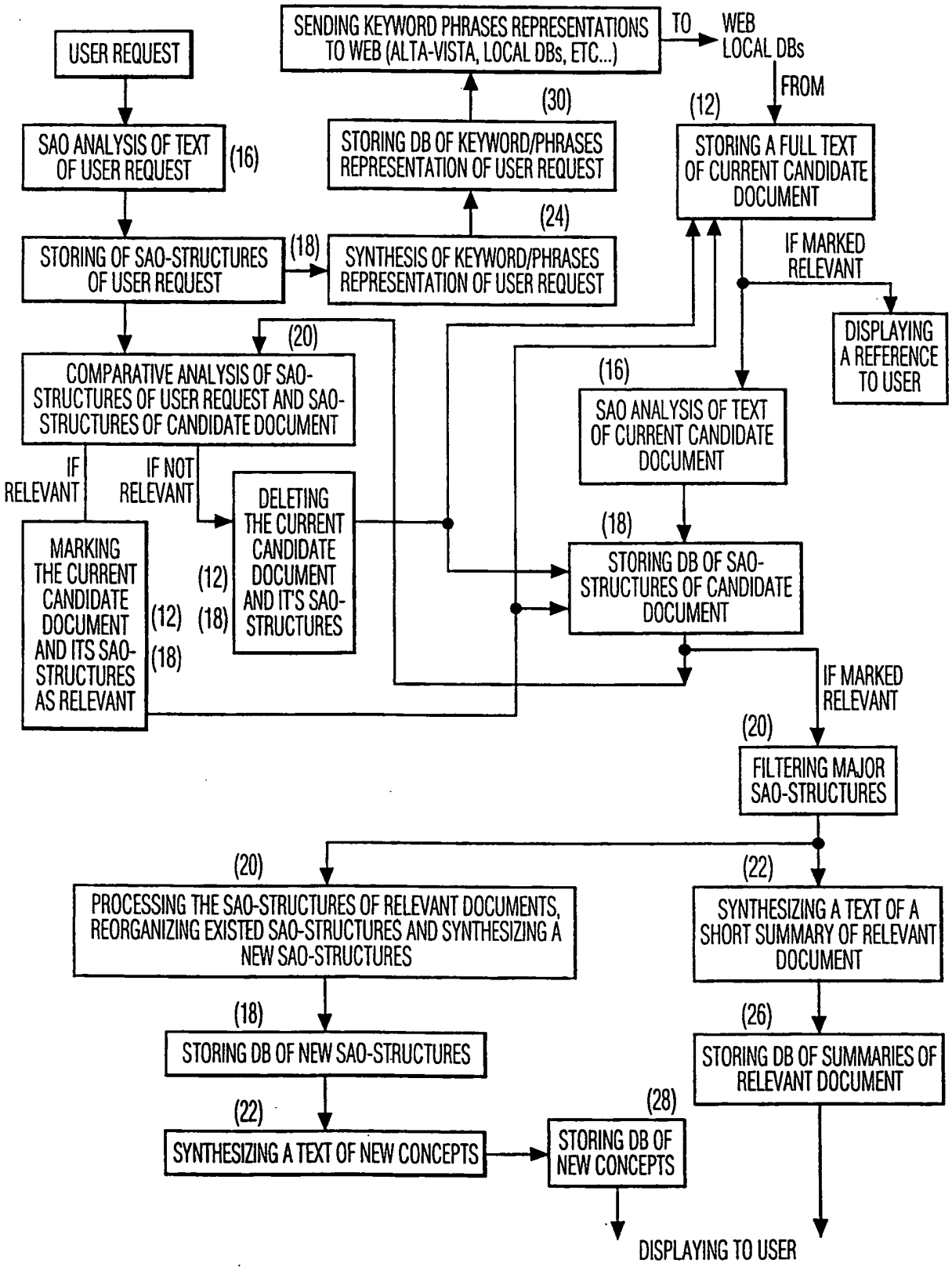

USER REQUEST
SAO ANALYSIS OF TEXT OF USER REQUEST
(16)
STORING OF SAO-STRUCTURES OF USER REQUEST
(18)
SYNTHESIS OF KEYWORD/PHRASES REPRESENTATION OF USER REQUEST
(24)
STORING DB OF KEYWORD/PHRASES REPRESENTATION OF USER REQUEST
(30)
SENDING KEYWORD PHRASES REPRESENTATIONS TO WEB (ALTA-VISTA, LOCAL DBs, ETC...)
TO WEB LOCAL DBs
FROM
(12)
STORING A FULL TEXT OF CURRENT CANDIDATE DOCUMENT
IF MARKED RELEVANT
DISPLAYING A REFERENCE TO USER
(20)
COMPARATIVE ANALYSIS OF SAO-STRUCTURES OF USER REQUEST AND SAO-STRUCTURES OF CANDIDATE DOCUMENT
IF RELEVANT
IF NOT RELEVANT
MARKING THE CURRENT CANDIDATE DOCUMENT AND ITS SAO-STRUCTURES AS RELEVANT
(12)
(18)
DELETING THE CURRENT CANDIDATE DOCUMENT AND IT'S SAO-STRUCTURES
(12)
(18)
(16)
SAO ANALYSIS OF TEXT OF CURRENT CANDIDATE DOCUMENT
(18)
STORING DB OF SAO-STRUCTURES OF CANDIDATE DOCUMENT
IF MARKED RELEVANT
(20)
FILTERING MAJOR SAO-STRUCTURES
(20)
PROCESSING THE SAO-STRUCTURES OF RELEVANT DOCUMENTS, REORGANIZING EXISTED SAO-STRUCTURES AND SYNTHESIZING A NEW SAO-STRUCTURES
(18)
STORING DB OF NEW SAO-STRUCTURES
(22)
SYNTHESIZING A TEXT OF NEW CONCEPTS
(28)
STORING DB OF NEW CONCEPTS
(22)
SYNTHESIZING A TEXT OF A SHORT SUMMARY OF RELEVANT DOCUMENT
(26)
STORING DB OF SUMMARIES OF RELEVANT DOCUMENT
DISPLAYING TO USER

FIG. 3

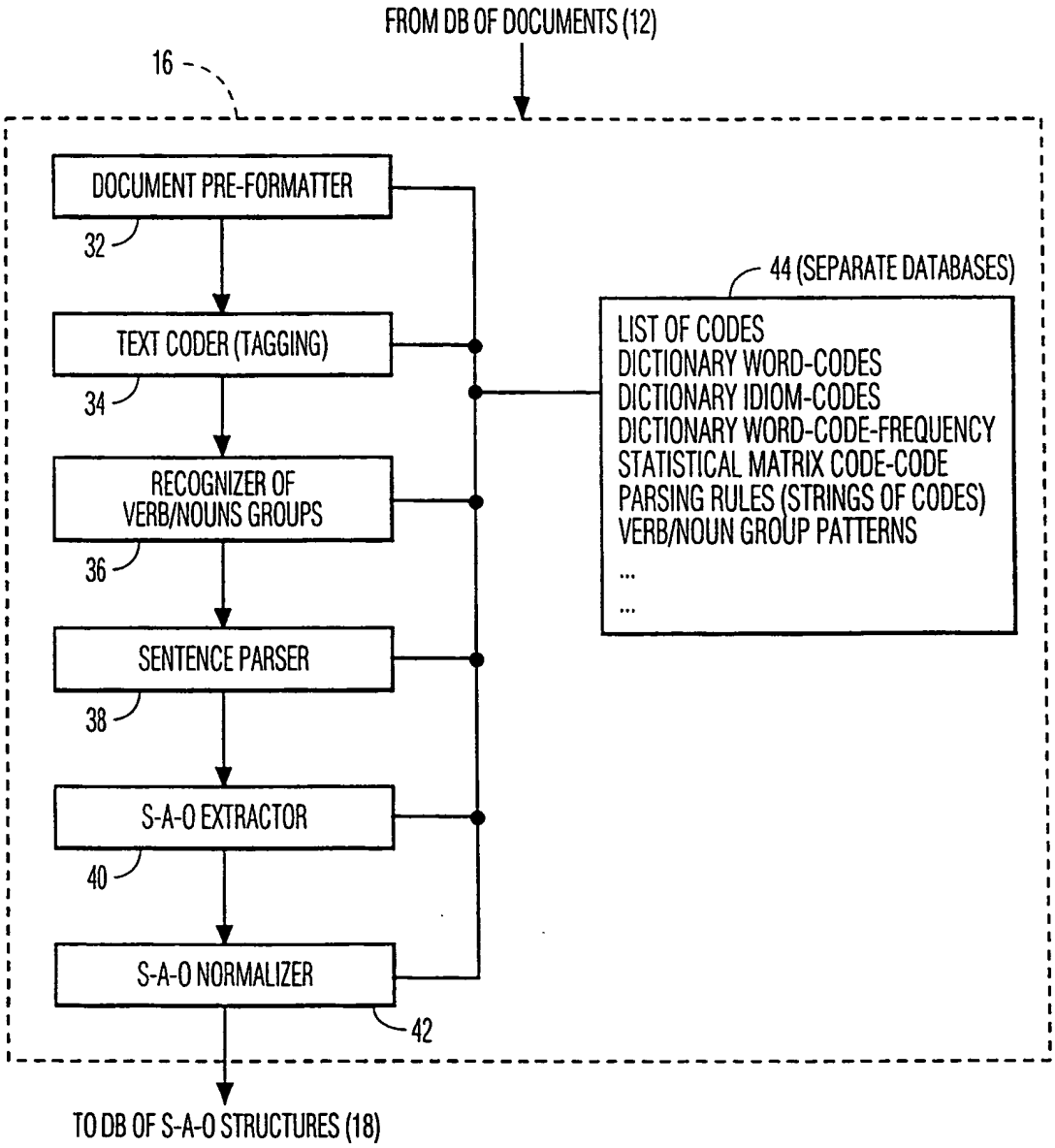

FROM DB OF DOCUMENTS (12)
16
DOCUMENT PRE-FORMATTER
32
TEXT CODER (TAGGING)
34
RECOGNIZER OF VERB/NOUNS GROUPS
36
SENTENCE PARSER
38
S-A-O EXTRACTOR
40
S-A-O NORMALIZER
42
44 (SEPARATE DATABASES)
LIST OF CODES
DICTIONARY WORD-CODES
DICTIONARY IDIOM-CODES
DICTIONARY WORD-CODE-FREQUENCY
STATISTICAL MATRIX CODE-CODE
PARSING RULES (STRINGS OF CODES)
VERB/NOUN GROUP PATTERNS
...
...
TO DB OF S-A-O STRUCTURES (18)

FIG. 4

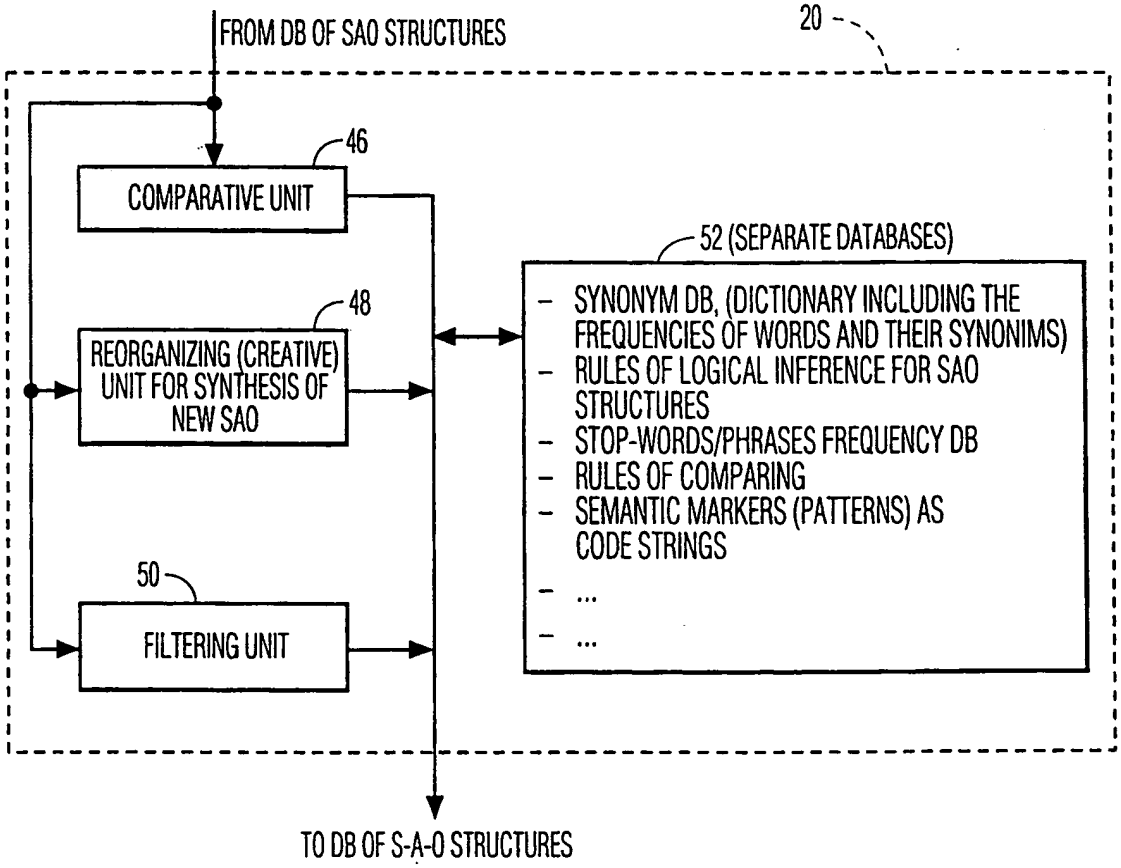

FROM DB OF SAO STRUCTURES
20
46
COMPARATIVE UNIT
52 (SEPARATE DATABASES)
48
REORGANIZING (CREATIVE) UNIT FOR SYNTHESIS OF NEW SAO
- SYNONYM DB, (DICTIONARY INCLUDING THE FREQUENCIES OF WORDS AND THEIR SYNONIMS)
- RULES OF LOGICAL INFERENCE FOR SAO STRUCTURES
- STOP-WORDS/PHRASES FREQUENCY DB
- RULES OF COMPARING
- SEMANTIC MARKERS (PATTERNS) AS CODE STRINGS
- ...
- ...
50
FILTERING UNIT
TO DB OF S-A-O STRUCTURES

FIG. 5

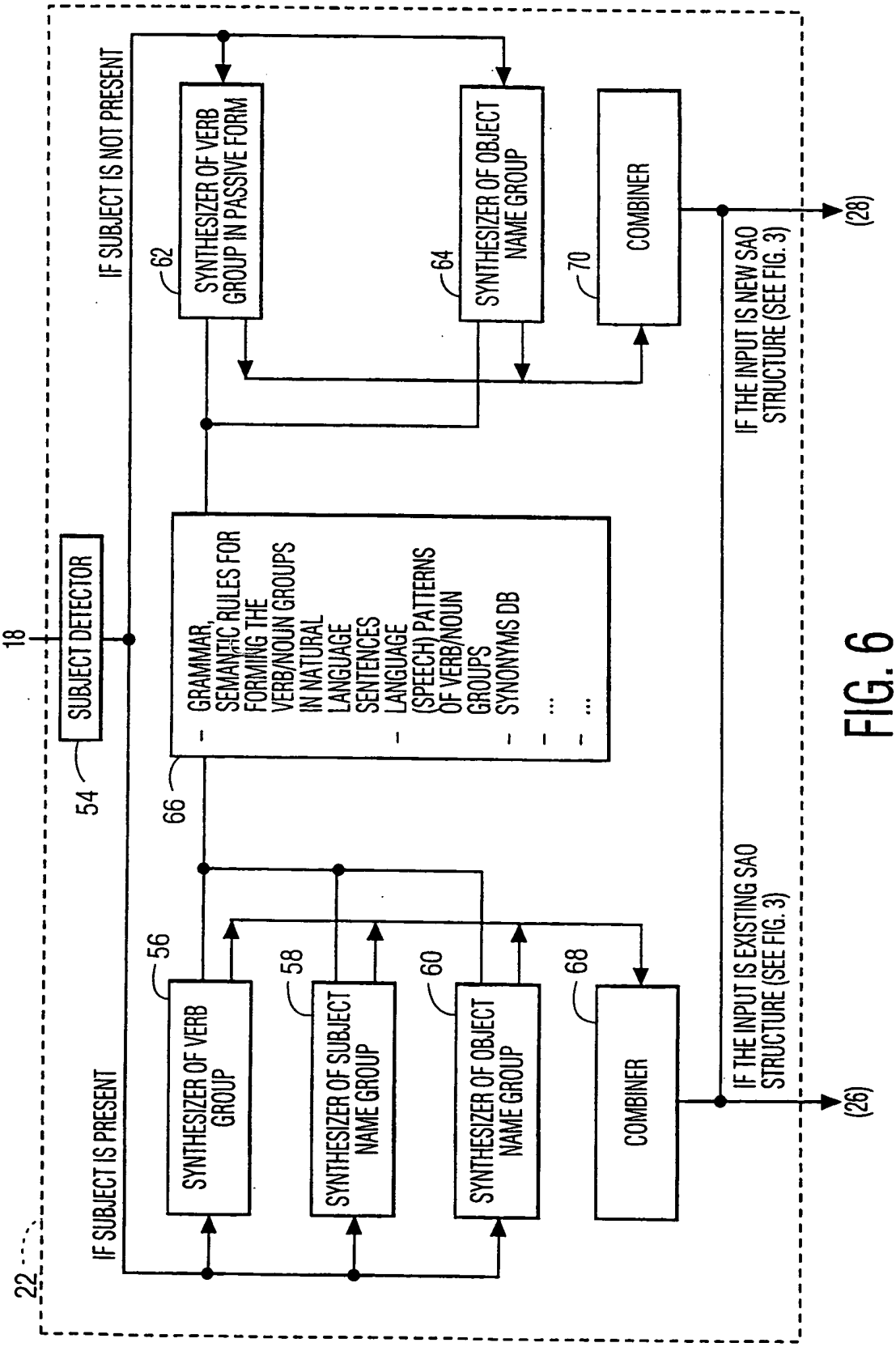

IF SUBJECT IS NOT PRESENT
SYNTHESIZER OF VERB GROUP IN PASSIVE FORM
62
SYNTHESIZER OF OBJECT NAME GROUP
64
COMBINER
70
IF THE INPUT IS NEW SAO STRUCTURE (SEE FIG. 3)
(28)
18
SUBJECT DETECTOR
54
- GRAMMAR, SEMANTIC RULES FOR FORMING THE VERB/NOUN GROUPS IN NATURAL LANGUAGE SENTENCES
- LANGUAGE (SPEECH) PATTERNS OF VERB/NOUN GROUPS
- SYNONYMS DB
- ...
- ...
66
IF SUBJECT IS PRESENT
56
SYNTHESIZER OF VERB GROUP
58
SYNTHESIZER OF SUBJECT NAME GROUP
60
SYNTHESIZER OF OBJECT NAME GROUP
68
COMBINER
IF THE INPUT IS EXISTING SAO STRUCTURE (SEE FIG. 3)
(26)
22

FIG. 6

SOURCE SENTENCE

The present invention shields a noise of an external magnetic field with the slider and improves a recording performance because the slider is isolated magnetically.

FIG. 7

TAGGED SENTENCE

The_ATI present_JJ invention_NN
shields_VBZ a_AT noise_NN of_IN
an_AT external_JJ magnetic_JJ
field_NN with_IN the_ATI slider_NN
and_CC improves_VBZ a_AT
recording_NN performance_NN
because_CS the_ATI slider_NN
is_BEZ isolated_VBN magnetically.

FIG. 8

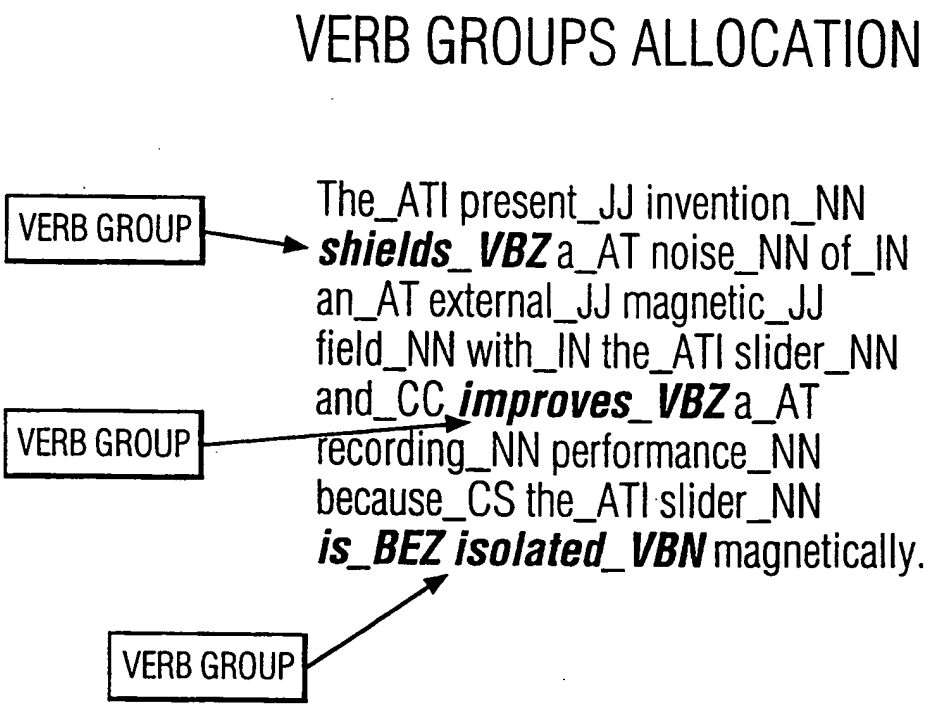

VERB GROUPS ALLOCATION
VERB GROUP
VERB GROUP
VERB GROUP
The_ATI present_JJ invention_NN
shields_VBZ a_AT noise_NN of_IN
an_AT external_JJ magnetic_JJ
field_NN with_IN the_ATI slider_NN
and_CC improves_VBZ a_AT
recording_NN performance_NN
because_CS the_ATI slider_NN
is_BEZ isolated_VBN magnetically.

FIG. 9

PCT/US99/19699

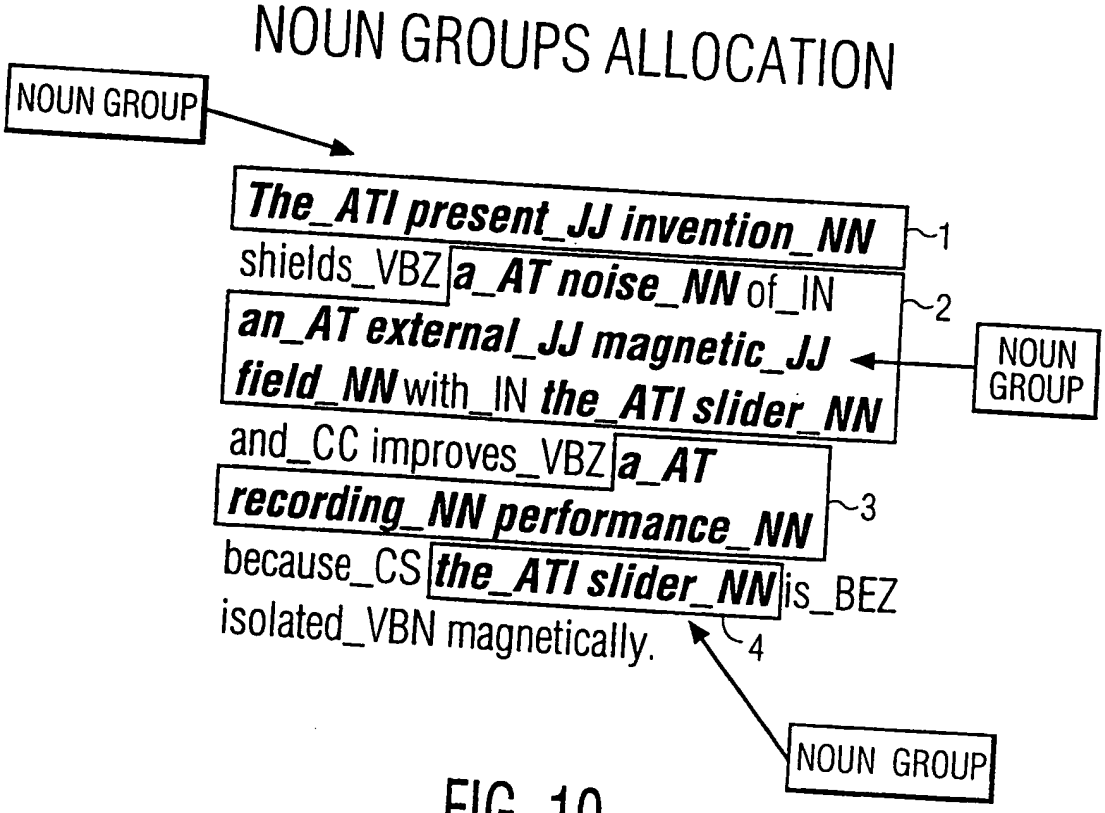


FIG. 10

SAO EXTRACTION

SUBJECT	ACTION	OBJECT
THE PRESENT INVENTION	SHIELDS	A NOISE OF EXTERNAL MAGNETIC FIELD
THE PRESENT INVENTION	IMPROVES	A RECORDING PERFORMANCE
	IS ISOLATED	THE SLIDER

FIG. 12

SAO EXTRACTION (NORMALIZED)

SUBJECT	ACTION	OBJECT
PRESENT INVENTION	SHIELD	NOISE OF EXTERNAL MAGNETIC FIELD
PRESENT INVENTION	IMPROVE	RECORDING PERFORMANCE
	ISOLATE	SLIDER

FIG. 13

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/19699

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 17/28, 21, 30
US CL :704/9

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 704/9, 1, 10; 707/2, 3, 4, 5, 6, 104, 530, 531, 532

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A, P	US 5,878,385 A (BRALICH, et al) 02 March 1999 abstract, fig. 1, col. 6, lines 13-60; col. 13, line 5 to col. 17, line 55	1-21
A, P	US 5,844,798 A (URAMOTO) 01 December 1998 abstract, col. 11, line 1 to col. 12, line 67; col. 16, line 1 to col. 19, line 11	1-21
A	US 5,802,504 A (SUDA et al) 01 September 1998 abstract, col. 1, line 16 to col. 2, line 33	1-21
X	US 5,799,268 A (BOGURAEV) 25 August 1998 abstract, figs. 1-5 & 7-14; col. 1, line 53 to col. 6, line 16; col. 6, line 65 to col. 12, line 27; col. 12, line 30 to col. 13, line 22; col. 39, line 40 to col. 42, line 51; col. 57, line 10 to col. 65, line 67	1-21

[X] Further documents are listed in the continuation of Box C. [] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

28 OCTOBER 1999

Date of mailing of the international search report

17 NOV 1999

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

JOSEPH THOMAS

Telephone No. (703) 308-3900

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

EAST

search terms: semantic processing, natural language, subject, action/verb, object, SAO/SVO, synonym, request, extract